

Análisis estadístico y representación gráfica con matri- ces en R



Nombre: Samuel Arteaga López
Curso: Sistemas de Big Data
Fecha de entrega: 31/01/2025

Índice

1. Creación de la matriz.....	3
2. Cálculos estadísticos.....	3
3. Representación gráfica.....	4
4. Preguntas de análisis.....	8
5. Entrega del código y resultados.....	8

En este ejercicio trabajarás con matrices en R para realizar cálculos estadísticos básicos y representarlos gráficamente. Sigue los pasos indicados y resuelve las preguntas relacionadas.

1. Creación de la matriz

- Genera una matriz de 5 filas y 10 columnas cuyos elementos sean números aleatorios entre 1 y 100. Asegúrate de que los valores sean reproducibles utilizando una semilla fija

```
set.seed(42)
print(matrix(sample(1:100, 50, replace = TRUE), nrow = 5, ncol = 10))
```

```
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
[1,]  49  18 100   3  95   3  30   8  92   99
[2,]  65  49  89  41   5  58  43  36  69   88
[3,]  25  47  37  89  84  97  15  68   4   87
[4,]  74  24  20  27  34  42  22  86  98   49
[5,] 100  71  26  36  92  24  58  18  50  26
```

Como se puede apreciar creamos la matriz con 5 filas y 10 columnas

2. Cálculos estadísticos

- Calcula los siguientes valores estadísticos para cada columna de la matriz:
 - Media
 - Mediana
 - Varianza
 - Desviación estándar

```
medias <- apply(matriz, 2, mean)
print("Media de cada columna:")
print(medias)
```

```
[1] 62.6 41.8 54.4 39.2 62.0 44.8 33.6 43.2 62.6 69.8 Media de cada columna
```

```
medianas <- apply(matriz, 2, median)
print("Mediana de cada columna:")
print(medianas)
```

```
[1] 65 47 37 36 84 42 30 36 69 87 Mediana de cada columna
```

```
varianzas <- apply(matriz, 2, var)
print("Varianza de cada columna:")
print(varianzas)
```

```
[1] 783.3 453.7 1392.3 988.2 1626.5 1271.7 294.3 1093.2 1437.8 957.7 Varianza
de cada columna
desviaciones <- apply(matriz, 2, sd) los parametros son, la matriz, columnas y fun-
cion
print("Desviación estándar de cada columna:")
print(desviaciones)

[1] 27.98750 21.30023 37.31354 31.43565 40.32989 35.66090 17.15517 33.06358
[9] 37.91833 30.94673 desviaciones de cada columna
```

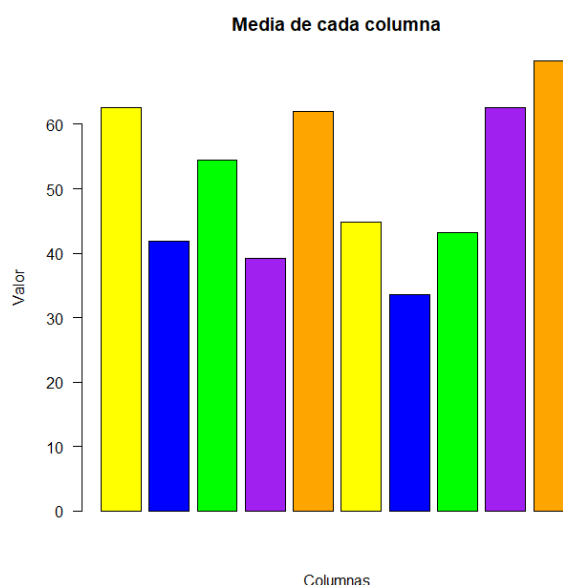
3. Representación gráfica

- Utiliza gráficos de barras para mostrar los valores calculados en el punto anterior.
- Cada gráfico debe contener:
- Un título descriptivo.
 - Ejes claramente etiquetados.
 - Barras con colores distintos para cada gráfico.

```
colores <- c("yellow", "blue", "green", "purple", "orange")
```

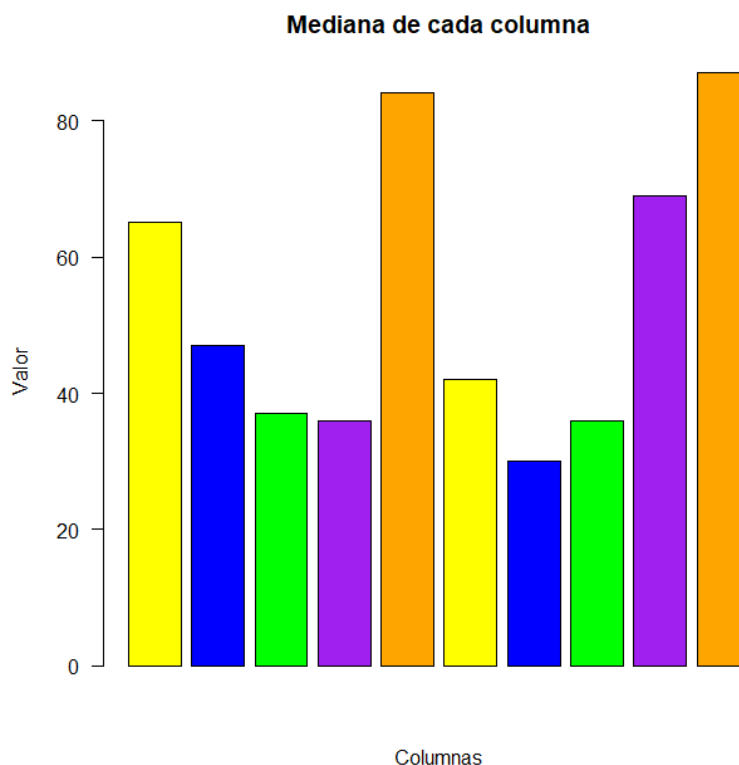
Para cada barplot tenemos como parámetros, la matriz, los colores, el título, los etiquetados en x e y, el color del borde y la orientación de las columnas

```
barplot(
  medias,
  col = colores,
  main = "Media de cada columna",
  xlab = "Columnas",
  ylab = "Valor",
  border = "black",
  las = 1
)
```



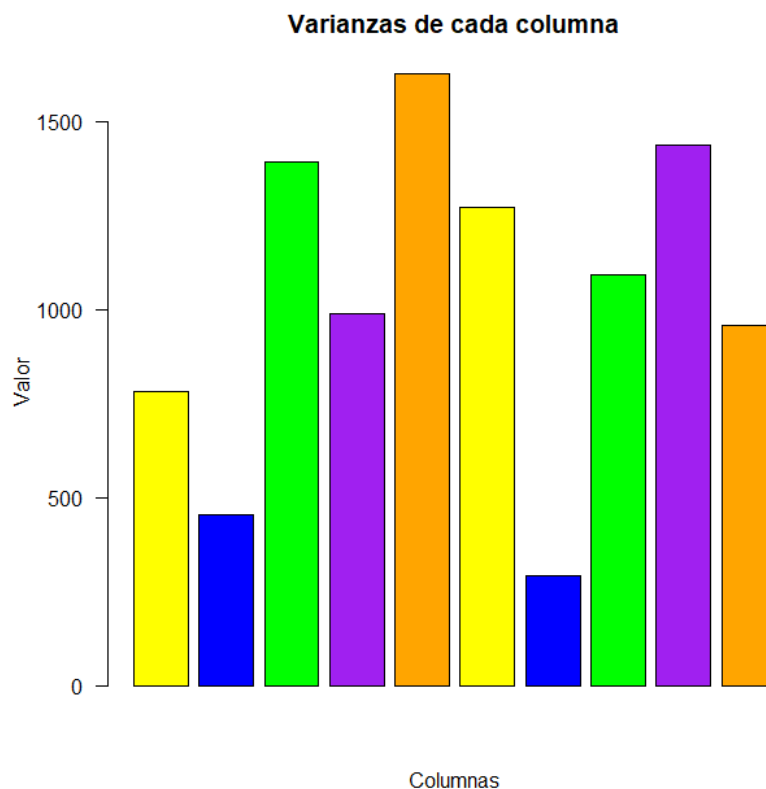
Breve descripción: Se puede apreciar una variedad de medias en cada columna, la columna 1, 2 y 9 tienen valores más altos que el resto de columnas, la más destacada es la 10, y la más baja la 7

```
barplot(  
  medianas,  
  col = colores,  
  main = "Mediana de cada columna",  
  xlab = "Columnas",  
  ylab = "Valor",  
  border = "black",  
  las = 1  
)
```



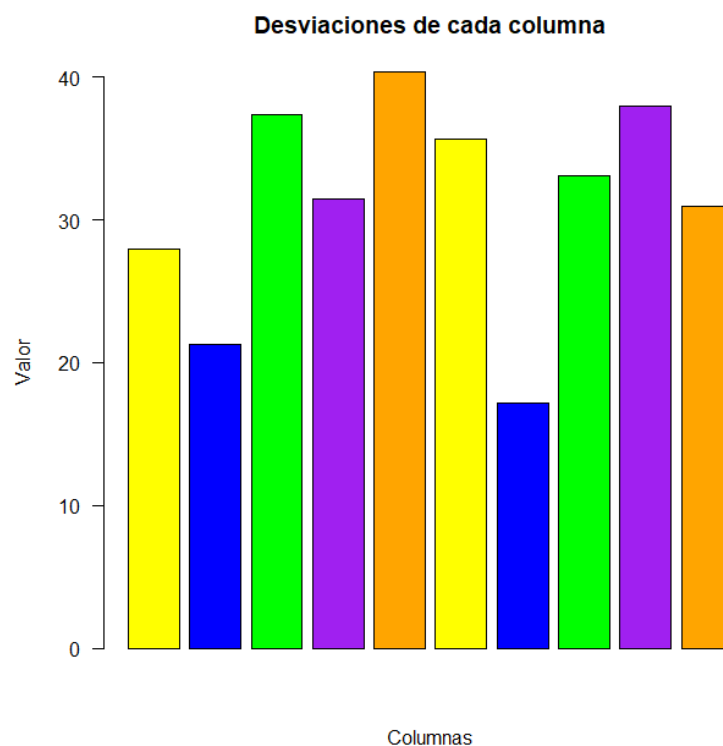
Breve descripción: Si la mediana de una columna es alta, esto indica que la mayoría de los valores en esa columna están distribuidos hacia el extremo superior de la escala. Si es baja, los valores de la columna tienden a estar distribuidos hacia el extremo inferior de la escala. Si la media y la mediana de una columna son similares, es probable que los datos estén distribuidos de manera relativamente simétrica. Si hay una diferencia significativa entre la media y la mediana, esto puede indicar que los datos están sesgados, es decir, que hay valores extremos que están tirando de la media hacia un lado.

```
barplot(
  varianzas,
  col = colores,
  main = "Varianzas de cada columna",
  xlab = "Columnas",
  ylab = "Valor",
  border = "black",
  las = 1
)
```

**Breve descripción:**

Si la varianza es alta, esto indica que los valores en la columna están muy dispersos alrededor de la media. Es decir, los datos tienen una mayor dispersión. Al contrario, si la varianza es baja, los valores están más concentrados alrededor de la media.

```
barplot(  
  desviaciones,  
  col = colores,  
  main = "Desviaciones de cada columna",  
  xlab = "Columnas",  
  ylab = "Valor",  
  border = "black",  
  las = 1  
)
```



Breve descripción: Se puede comprobar mediante la altitud de las barras de las columnas que los datos están dispersos en cada columna, si estuvieran igualitarias las barras, no habría dispersión de los datos, serían casi todos iguales.

4. Preguntas de análisis

Contesta las siguientes preguntas basándote en los resultados obtenidos:

- ¿Qué columna tiene la media más alta? [La columna 10](#)
- ¿Qué columna tiene la varianza más baja? [La columna 7](#)
- ¿Existe alguna columna en la que la mediana sea significativamente distinta de la media?
[Si la columna 7](#)
- ¿Cómo describirías el comportamiento general de la desviación estándar entre las columnas? [El comportamiento general se describe como dispersos, hay una mayor variabilidad en los datos](#)

5. Entrega del código y resultados

- Copia el código utilizado para generar la matriz, realizar los cálculos y producir las gráficas.
- Incluye una breve interpretación de los resultados obtenidos en los gráficos y responde las preguntas de análisis.

Pista: Puedes usar las funciones `matrix()`, `apply()`, y `barplot()` en R para completar el ejercicio.